# UTA100 2023 比赛回顾

#### Jun Zhang

## 前言

UTA100 2023已于数月前落幕，今年自己没有上场比赛只是以后援的身份参与其中，也是一份难得的体验。趁着最近有时间，就从统计的角度来回顾一下这场比赛。

2019年之前的UTA比赛数据，官方都提供Excel文件可以直接下载。自从UTA换了东家之后，现在必需自己想办法从网上抓取。本短文中所使用的比赛数据均来自于UTA2023官方所使用的Multi Sport Australia网站 ( <https://www.multisportaustralia.com.au/races/ultra-trail-australia-2023/> )。

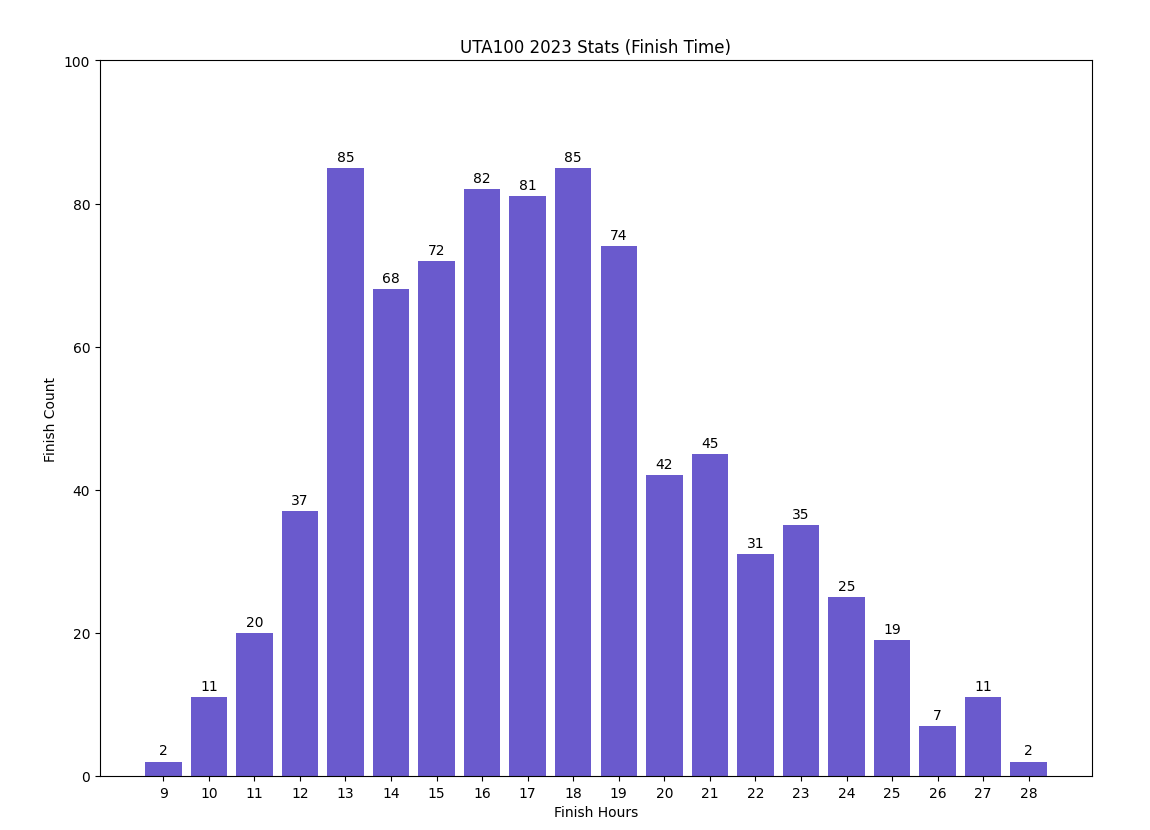
## 比赛统计信息

### 概要信息

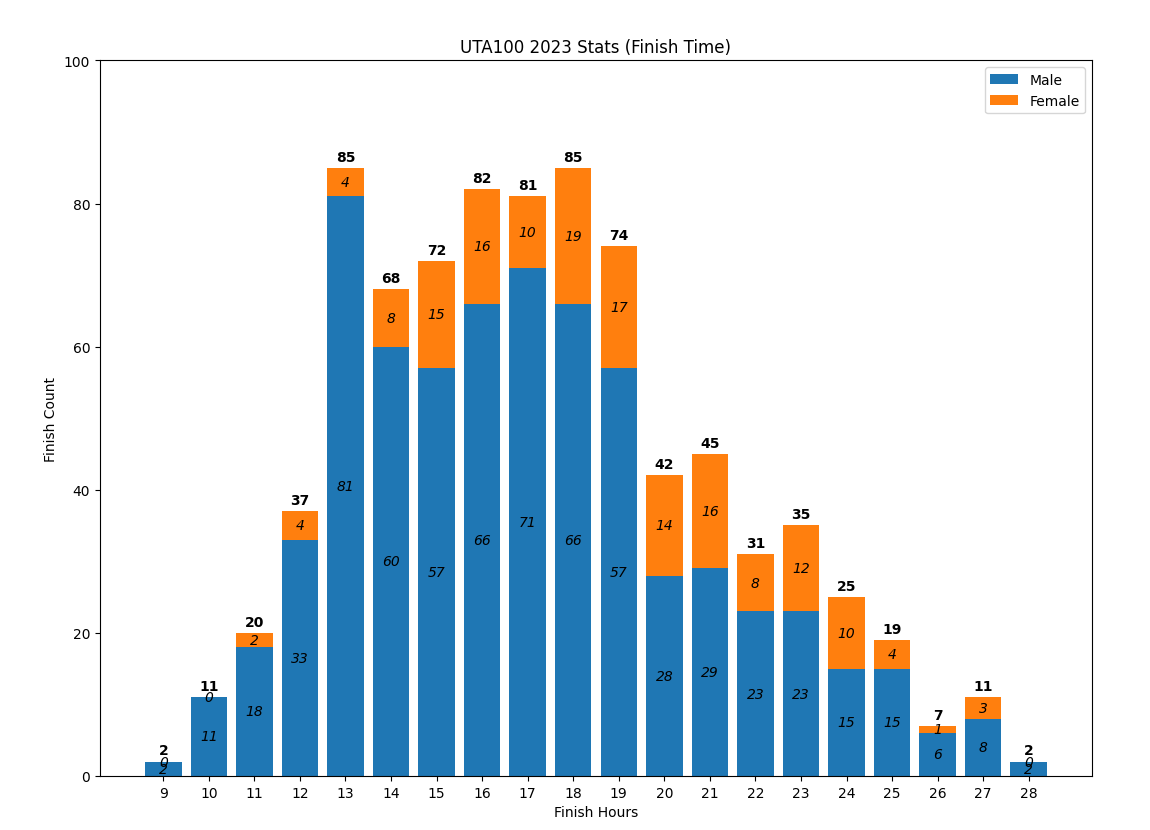
今年的UTA100重回以前的经典比赛路线，总共有1108人报名参赛，其中完赛834人，未完成149人，未出发125人。

|  |  |
| --- | --- |
| **Status** | **Total** |
| Finished | 834 |
| Did Not Finish | 149 |
| Not Yet Started | 125 |

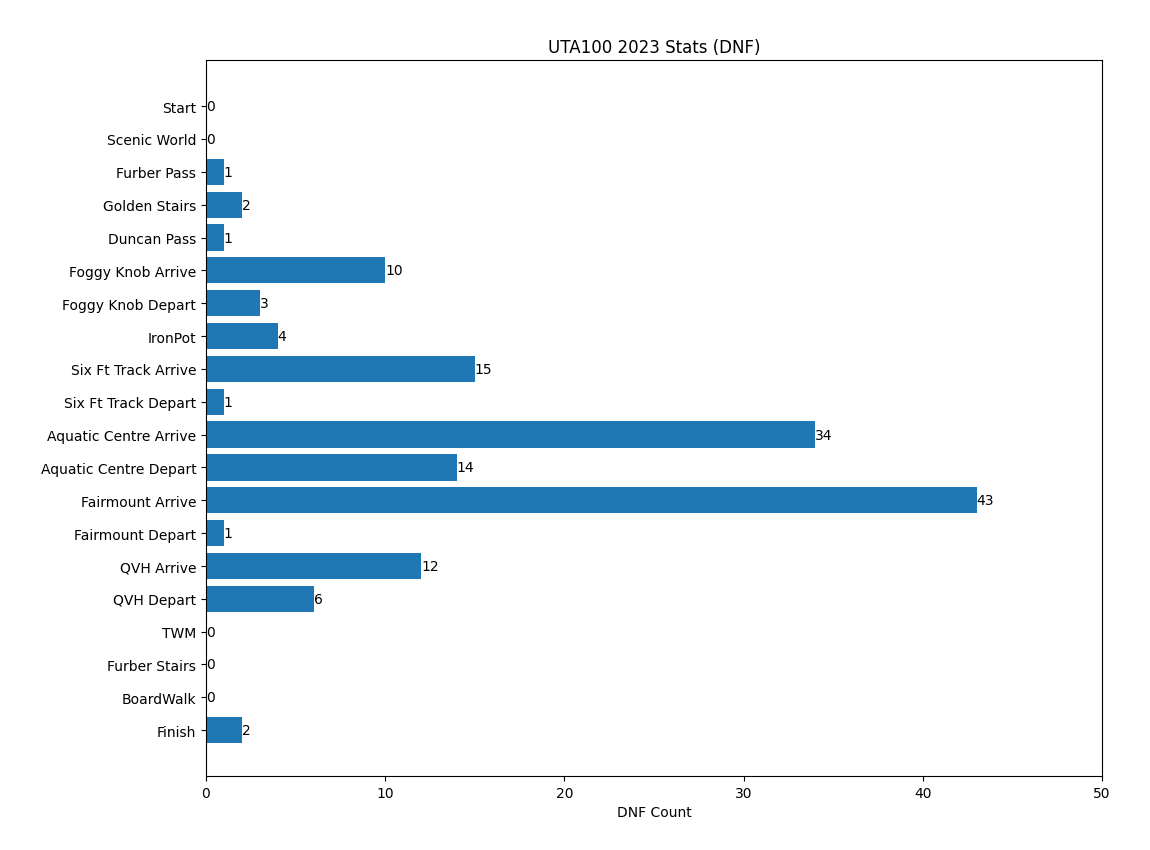
### **完赛时间分布**



从图中可以看出，12h以内完赛的只有33人，越野精英人群；13h到14h以内这一时段的人数很多，这些选手就是奔着银扣去的，够拼够努力；中间14h至20h以内完赛的人数居多，占了大多数；随着比赛的延续，再往后完赛的人数逐渐减少。

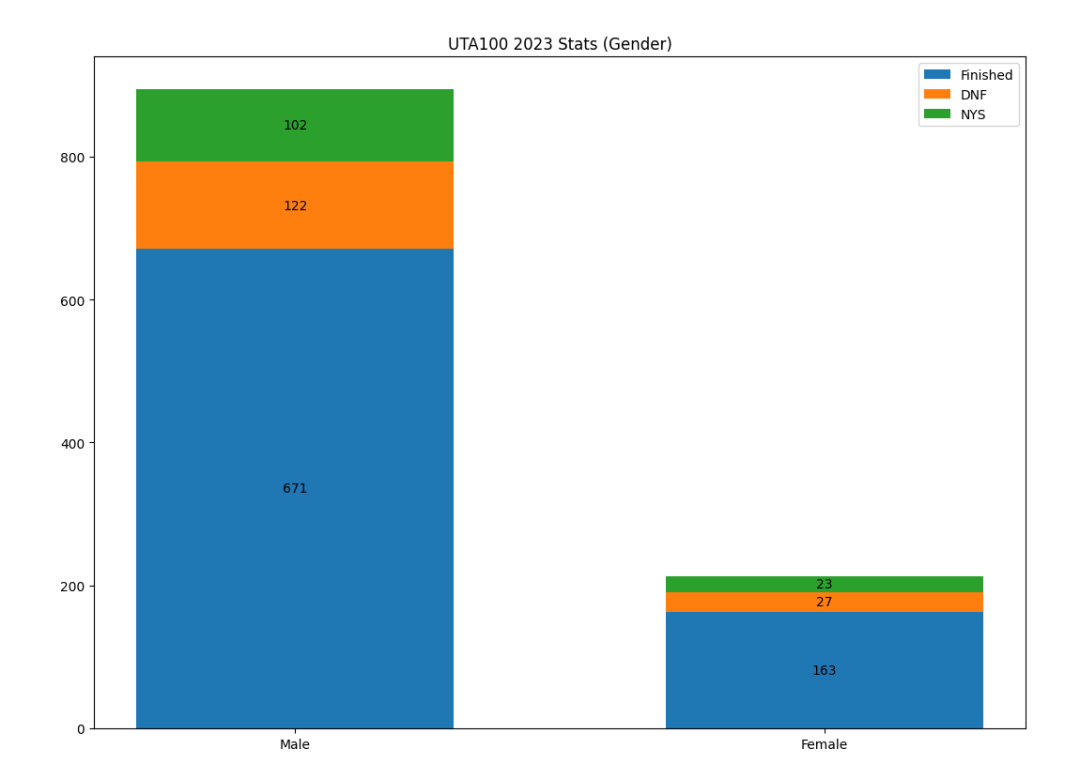


### **未完赛(DNF)**



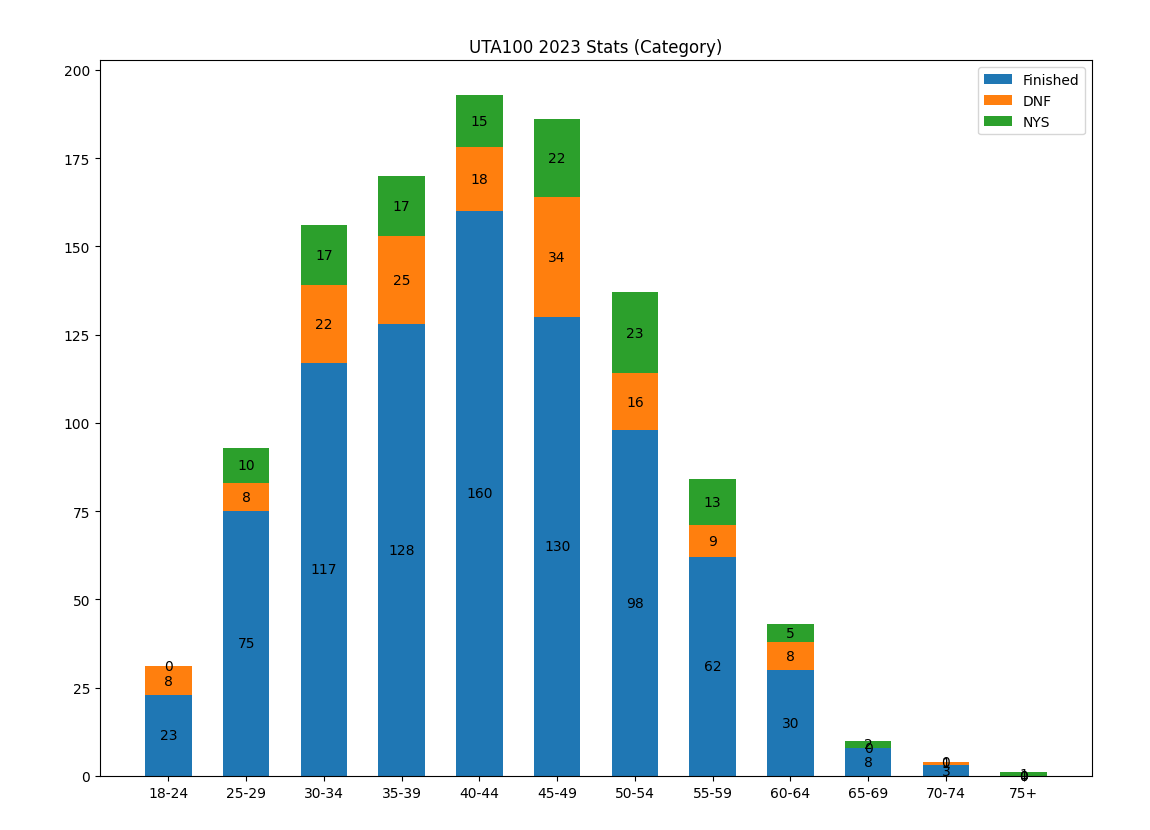
这个图中统计的是DNF选手最后计时点的位置，以进入CP5和CP4之后退赛的居多，占总DNF人数的51.7%。

### **性别分组**



其中，男子报名895人，完赛671人，完赛率75.0%；女子报名213人，完赛163人，完赛率76.5%。

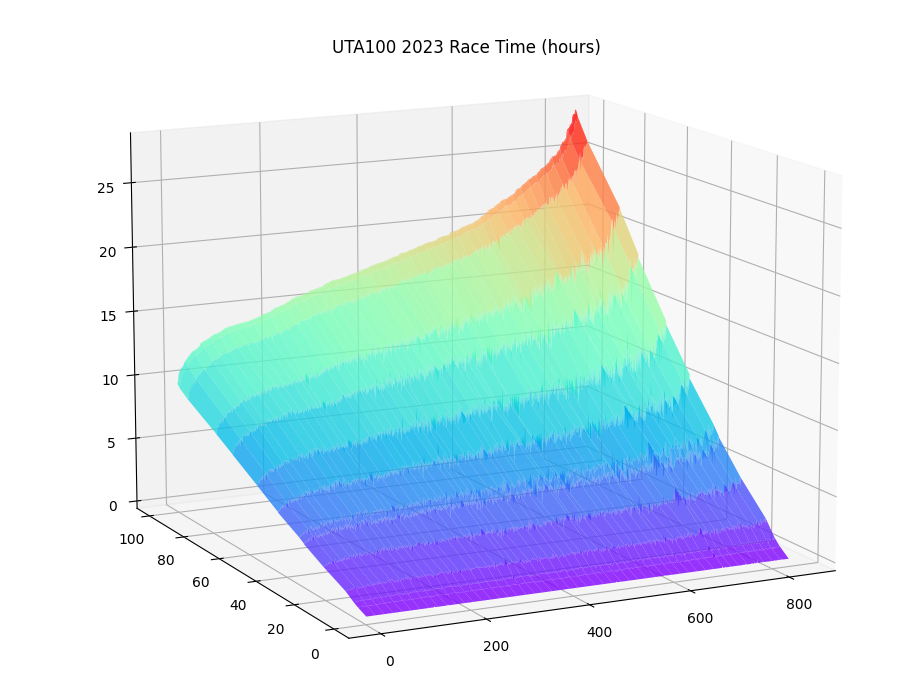
### **年龄组别**

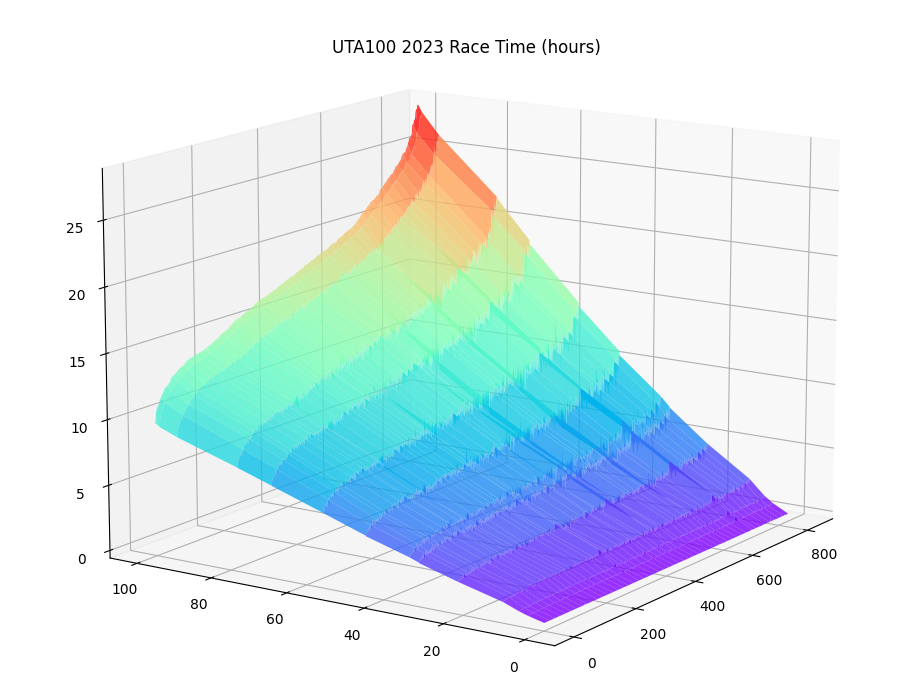


可以看出今年UTA100的三大主要年龄组分别为40-44， 45-59和35-39，占总报名人数的49.5%，总完赛的50.1%，而年龄最大的70-74组亦有4人报名3人完赛。

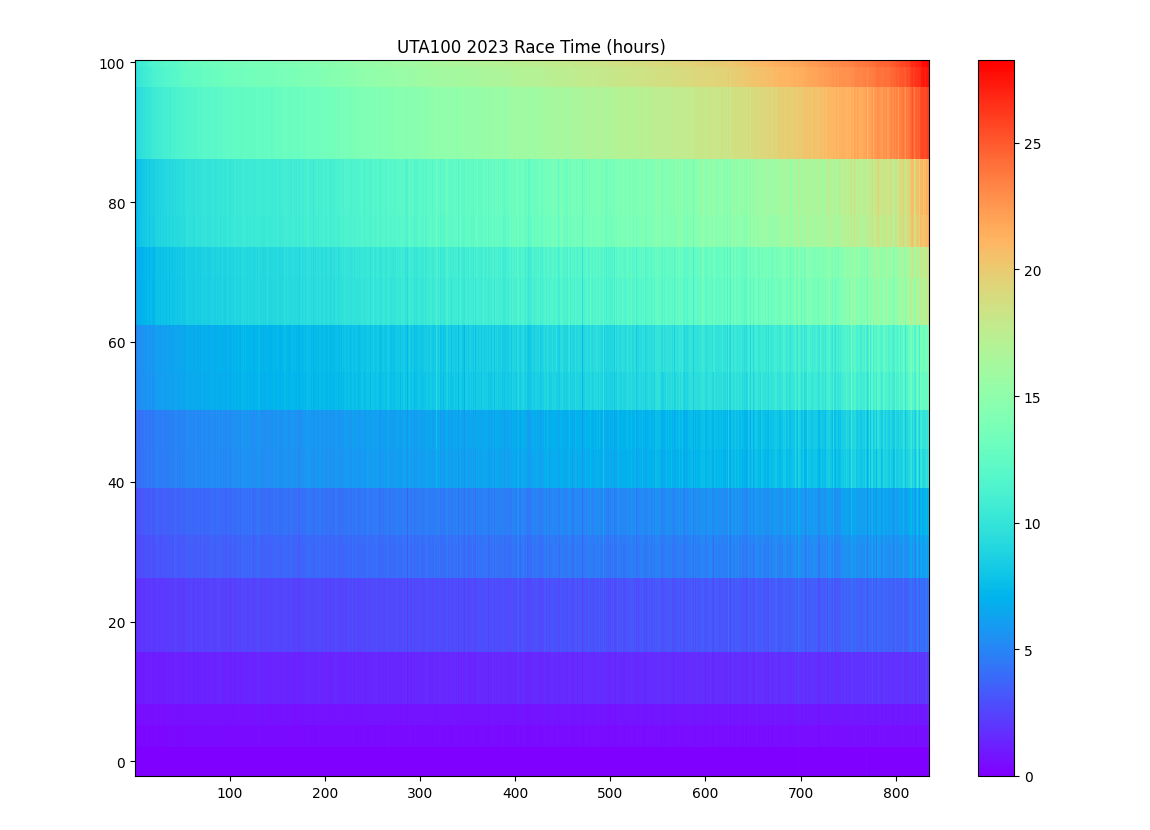
## 比赛用时分析

### **比赛时序视图**

****

****

这个3D视图中，横向是比赛名次，纵向是里程，竖向是完赛时间(hours)。从图中可以看出，前面的精英选手和后面的差距明显，而20小时之后，完赛时间增加显著，长时间的运动让完赛变得越来越困难、越拖越漫长，中段12h到20h完赛的数据相对线性、平稳。

****

这是投影到平面上的2D视图。

### **比赛时序分析**

定义比赛中间计时点*n*的分段用时比值*Pn* (*Proportion*)为：

这里*n = 2 ... 19*。该数据项的意思是指第*n*个计时点到前一个计时点*n-1*的用时*Tn-1, n*，与后一个计时点*n+1*到前一个计时点*n-1*的用时*Tn-1, n+1*之间的比值。

在抓取到原始比赛结果后，用所有完赛选手的数据，计算出各计时点之间的分段用时比值为：

|  |  |  |
| --- | --- | --- |
| **Location (n)** | **Proportion (P)** | **Standard Deviation (σ)** |
| 2 | 0.586274 | 0.012317 |
| 3 | 0.290029 | 0.012318 |
| 4 | 0.348253 | 0.025149 |
| 5 | 0.495898 | 0.024340 |
| 6 | 0.937742 | 0.038332 |
| 7 | 0.163351 | 0.089850 |
| 8 | 0.219899 | 0.024682 |
| 9 | 0.909416 | 0.049517 |
| 10 | 0.082487 | 0.045936 |
| 11 | 0.883966 | 0.068089 |
| 12 | 0.093464 | 0.055632 |
| 13 | 0.926608 | 0.049498 |
| 14 | 0.100476 | 0.065256 |
| 15 | 0.921762 | 0.066937 |
| 16 | 0.048146 | 0.042810 |
| 17 | 0.723417 | 0.022075 |
| 18 | 0.750221 | 0.023933 |
| 19 | 0.914246 | 0.013235 |

尽管不同选手之间的完赛时间各不相同，从9.5小时到28小时，可以说相差很大，但透过前面的比赛时序视图可以看到，完赛选手总体上在时间分配比赛进程上大都还是比较平稳的。

从上面这个表中也可以看到，在UTA100比赛数据中，分段用时比值*Pn*这个量的标准差较小，意味着其离散程度较小，可以利用它来进行数据检查修复，以及建立比赛用时规划模型。

### **粗差检查和数据修复**

分段用时比值数据项首先用作粗差检测，基于完赛选手的比赛数据计算单个*Proportion*和对应均值之差，用六倍标准差**(±6σ)**作为显著性粗差判断标准，对超出这个范围的记录进行人工检查。基本原则是：

1. 一般路段的分段用时如果过短，则计时有误；
2. CP点停留时间可以较长；
3. 其它情况则保留原始记录不修正。

依据上述标准，共有9个记录被标记为粗差。经过人工检查之后，确认并修正了5个记录，保留了另外4个记录。

接着用它来进行数据修复的工作，这也正是最初引入该数据项的直接原因。在官方公布的比赛数据中，因为各种原因缺少了23条计时记录，利用前后计时时间差*Tn-1, n+1*和相应的分段用时比值*Pn*，计算出*Tn-1, n*以补齐这些缺失的数据。

和*Race Time*一并补齐的还有*Time of Day*和*Split Time*这两项。

在完成了数据记录修正和补齐之后，计算出最后的分段用时比值：

|  |  |  |
| --- | --- | --- |
| **Location (n)** | **Proportion (P)** | **Standard Deviation (σ)** |
| 2 | 0.586274 | 0.012288 |
| 3 | 0.290013 | 0.012294 |
| 4 | 0.347259 | 0.015259 |
| 5 | 0.497011 | 0.013817 |
| 6 | 0.937767 | 0.038335 |
| 7 | 0.163015 | 0.088253 |
| 8 | 0.220800 | 0.016568 |
| 9 | 0.909359 | 0.049429 |
| 10 | 0.082487 | 0.045936 |
| 11 | 0.883966 | 0.068089 |
| 12 | 0.093423 | 0.055612 |
| 13 | 0.926632 | 0.049327 |
| 14 | 0.100476 | 0.065021 |
| 15 | 0.921674 | 0.066785 |
| 16 | 0.048146 | 0.042759 |
| 17 | 0.723415 | 0.022050 |
| 18 | 0.750221 | 0.023933 |
| 19 | 0.914246 | 0.013235 |

### **分段用时与比赛用时**

假如我们把第一个路段用时标记为*T12*，应用上面计算得到的分段用时比值*P2*可以计算出第二个路段用时*T23*为：

之后的*T34*则为：

那么*Tn, n+1*就是：

即为：

计算出各个分时段的用时之后，总的比赛用时就表达为：

### **比赛用时规划**

总结过去是为了展望未来，利用过往比赛数据，可以帮助我们更好地准备以后的比赛。当进行比赛用时规划时，先设定预计的完赛目标时间*Trace*，通过反向计算得到未知的*T12*项：

接着依次计算出后续各个路段用时*Tn, n+1*：

将这些计算出来的路段用时串连起来，再考虑比赛分组出发时间，最终就得到一份完整、可行且稳健的比赛用时计划。

## 小结

本短文简要回顾2023年的UTA100比赛数据，并据此修正补齐数据集，并建立比赛用时规划模型，为以后参加UTA100比赛提供参考。这里介绍的概念与方法已被应用在[UTA100 Planner](https://utaplanner.vercel.app/) (<https://utaplanner.vercel.app/>) 这个App中，其中所用到的模型参数也已更新至最新，后续还将进行进一步优化。

这里提出的规划方法不只是针对UTA100这一项比赛，也可以套用在其它赛事准备上，当然前提条件是要有过往比赛足够多的中间计时点数据，且赛道线路前后没有较大改动才行。象UTA50这样中间只有少数几个计时点的情况，勉强套用并没有参考意义。